Clean Copy

Appendix A

Partial Least Squares regression (PLSR)

Let a multivariate regression model be defined as:

$Y = XB + F$

where
X a *NxP* matrix with *N* predictor variables (genes);
Y (*NxJ*) being the *J* predicted variables. In our case Y represents a matrix containing dummy variables;
B is a matrix of regression coefficients; and
F is a *NxJ* matrix of residuals.

The structure of the PLSR model can be written as:

$X = TP^T + E_A$, and
$Y = TQ^T + F_A$, where

where
T (*NxA*) is a matrix of score vectors which are linear combinations of the x-variables;
P (*PxA*) is a matrix with the x-loading vectors p_a as columns;
Q (*JxA*) is a matrix with the y-loading vectors q_a as columns;
E_a (*NxP*) is the matrix for X after A factors; and
F_a (*NxJ*) is the matrix for Y after A factors.

The criterion in PLSR is to maximize the explained covariance of [X,Y]. This is achieved by the loading weights vector w_{a+1}, which is the first eigenvector of $E_a^T F_a F_a^T E_a$ (E_a and F_a are the deflated X and Y after *a* factors or PLS components).

The regression coefficients are given by:

$B = W(P^TW)^{-1}Q^T$

A PLSR model with full rank, i.e. maximum number of components, is equivalent to the MLR solutions. Further details on PLSR can be found in Marteus & Naes, 1989, Multivariate Calibration, John Wiley & Sons, Inc., USA and Kowalski & Seasholtz, 1991, supra.

Clean Copy

Example 3: Validation of Example 1, diagnosis of breast cancer

The results in Example 1 were validated by using the informative probes identified in Example 1 on new beast cancer and control samples.

Methods

The methods, essentially as described in Example 1, were used. Blood was taken from patients as described in Table 8. However, blood was collected in PAXgene tubes and the first strand labelled cDNAs were hybridized to 719 cDNAs spotted on nylon membranes along with other controls as described in Example 1. After background subtraction using control spots, the data of each membrane was normalized using the inter quantile range. The data was analysed as described in Example 1 and the model validated by cross validation.

The 719 cDNAs which were spotted are a subset of the cDNAs spotted in Example 1 and include 111 cDNAs described in Table 2 and which were found to be informative in Example 1.

Results

The results are shown in Figures 4 to 9. Figures 4, 6 and 8 are projection plots similar to Figure 2 and show the projection of normal and breast cancer patients' samples onto a classification model generated using all 719 cDNA. Figure 6 is similar but uses a classification model generated with the 111 probes common to Example 1. Figure 8 uses the 345 sequences of the 719 for which sequence information is provided herein. In each case classification of normal and breast cancer groups was possible. Figures 5, 7 and 9 show prediction plots which reflect the ability of the generated models to

correctly diagnose breast cancer. In the 3 prediction plots shown, the disease samples appear on the x axis at +1 and the non-disease samples appear at -1. The y axis represents the predicted class membership. During prediction, if the prediction is correct, disease samples should fall above zero and non-disease samples should fall below zero. In each case almost all samples are correctly predicted.

Example 4: Validation of Example 2, diagnosis of Alzheimers

The results in Example 2 were validated by using the informative probes identified in Example 2 on new Alzheimer's patient samples.

Methods

The methods, essentially as described in Example 2, were used. Twelve female patients diagnosed with Alzheimer's disease at the Memory Clinic at Ullevål University Hospital who were confirmed as having Alzheimer's disease based on the criteria of Example 2 were used in the trial. The mean age of the patients was 72.3 with an age range of 66-83. The mean MMSE score was 22.0 (the maximum score attainable being 30).

Sixteen age-matched female individuals without diagnosed Alzheimer's disease were used as the normal control group. All had been tested with MMSE and had a minimum score of 29. The mean age of the normal control group was 74.0 and the age range 66-86.

After transfer of the blood to PAXgene tubes, total mRNA was isolated from the blood of the Alzheimer's disease and from the control group donors according to the manufacturers's instructions (PreAnalytiX,

Clean Copy

Hombrechtikon, Switzerland). The isolated mRNA was labelled during reverse transcription in the presence of α^{33}P-dATP, yielding a labelled first strand cDNA. Hybridization was performed as described previously onto 730 cDNA clones picked from a cDNA library from whole blood of 550 healthy individuals without knowledge of the gene sequence of the random cDNA clones.

Results

The results are shown in Figures 10 and 11. Figure 10 is a projection plot generated using 520 probes which have been sequenced. Figure 11 is a prediction plot and shows correct prediction of almost all samples.

Clean Copy

Table 1a

List of probes informative for disease diagnosis

	Clone ID	No. of nucleotides	SEQ ID NO: in sequence listing
1	I-24	373	11
2	I-28	564	13
3	I-30	622	398
4	I-34	554	15
5	I-54	156	399
6	I-58	554	24
7	II-03	622	34
8	II-05	628	35
9	II-06	528	36
10	II-10	329	39
11	II-24	534	47
12	II-25	444	48
13	II-26	566	49
14	II-33	523	55
15	II-34	566	56
16	II-41	534	60
17	II-42	512	61
18	II-57	505	73
19	II-61	596	77
20	II-69	387	85
21	II-70	420	86
22	II-75	535	91
23	II-84	577	99
24	II-87	552	100
25	II-88	606	101
26	II-94	329	104
27	III-02	747	107
28	III-06	682	109

29	III-08	536	111
30	III-13	615	115
31	III-20	479	401
32	III-23	694	119
33	III-26	476	122
34	III-35	551	130
35	III-39	224	131
36	III-40	349	132
37	III-43	382	500
38	III-44	382	134
39	III-53	390	142
40	III-56	109	144
41	III-57	374	145
42	III-61	521	148
43	III-63	575	150
44	III-74	502	155
45	III-80	585	158
46	III-85	516	161
47	III-89	660	165
48	IV-14	545	275
49	IV-15	628	402
50	IV-26	494	403
51	IV-31	268	278
52	IV-32	569	279
53	IV-53	362	498
54	IV-69	286	4
55	IV-80	579	291
56	IX-10	641	314
57	IX-38	583	317
58	IX-39	424	318
59	IX-48	626	319
60	IX-77	556	325
61	V-03	496	296

Clean Copy

62	V-04	397	297
63	V-07	293	298
64	V-11	599	404
65	V-12	498	301
66	V-55	421	499
67	V-80	260	311
68	VI-04	122	339
69	VI-07	405	1
70	VI-12	667	341
71	VI-14	642	343
72	VI-20	115	346
73	VI-23	634	347
74	VI-48	626	355
75	VI-50	585	356
76	VI-53	560	357
77	VI-55	509	359
78	VI-70	550	2
79	VI-74	655	365
80	VI-76	582	367
81	VI-87	595	370
82	VI-88	651	371
83	VI-95	230	374
84	VII-03	412	411
85	VII-15	439	414
86	VII-19	580	171
87	VII-21	671	173
88	VII-32	457	179
89	VII-36	209	182
90	VII-39	541	183
91	VII-42	502	186
92	VII-43	316	187
93	VII-46	631	190
94	VII-47	526	415

Clean Copy

95	VII-48	613	416
96	VII-59	565	199
97	VII-63	98	201
98	VII-66	362	204
99	VII-72	595	206
100	VII-73	522	207
101	VII-76	624	209
102	VII-77	692	418
103	VII-80	338	210
104	VII-81	556	211
105	VII-90	576	216
106	VII-91	341	217
107	VII-93	379	219
108	VIII-09	598	221
109	VIII-20	419	229
110	VIII-28	511	235
111	VIII-29	592	236
112	VIII-30	572	237
113	VIII-31	482	238
114	VIII-32	545	239
115	VIII-33	624	240
116	VIII-41	649	245
117	VIII-42	600	246
118	VIII-46	425	249
119	VIII-48	251	251
120	VIII-64	627	261
121	VIII-66	345	262
122	VIII-67	252	263
123	VIII-76	691	270
124	X-07	641	328
125	X-15	132	329
126	X-29	370	331

127	X-54	603	334
128	X-56	71	335
129	X-68	642	421
130	X-72	622	336
131	X-94	601	337
132	XI-13	620	423
133	XI-81	374	426
134	XII-07	567	427
135	XII-35	620	428
136	XII-59	484	430
137	XIII-19	559	433
138	XIII-52	513	378
139	XIII-92	741	435
140	XV-22	-	388
141	XV-25	485	436
142	XVI-36	435	382
143	XVI-53	741	439
144	XVI-66	689	384
145	XVI-76	198	386
146	XVI-77	198	387
147	XVII-31	503	392
148	XVII-40	203	440
149	XVII-48	587	393
150	XVII-76	650	394
151	XVII-87	502	395
152	XVII-95	648	396

Clean Copy

Table 1b

List of sequences of probes informative for disease diagnosis

Clone ID	SEQ ID NO. in Sequence Listing
I-10	6
I-13	444
I-14	397
I-15	7
I-17	8
I-19	9
I-22	10
I-24	11
I-25	12
I-28	13
I-30	398
I-31	14
I-34	15
I-37	482
I-38	16
I-39	17
I-40	18
I-42	445
I-48	19
I-49	20
I-53	21
I-54	399
I-56	22
I-57	23
I-58	24
I-60	25
I-64	26
I-67	27
I-69	28
I-77	29
I-80	30
I-81	31
I-82	32
I-86	447
I-88	400
I-95	448
II-02	33
II-03	34
II-05	35
II-06	36
II-07	37
II-08	38
II-10	39
II-11	40
II-12	41
II-13	42
II-15	43
II-16	44
II-21	45
II-23	46
II-24	47
II-25	48
II-26	49
II-27	50
II-29	51
II-30	52
II-31	53
II-32	54
II-33	55
II-34	56
II-38	57
II-39	58
II-40	59
II-41	60

Clean Copy

II-42	61
II-43	62
II-44	63
II-46	64
II-47	65
II-48	66
II-50	67
II-52	68
II-53	69
II-54	70
II-55	71
II-56	72
II-57	73
II-58	74
II-59	75
II-60	76
II-61	77
II-62	78
II-63	79
II-64	80
II-65	81
II-66	82
II-67	83
II-68	84
II-69	85
II-70	86
II-71	87
II-72	88
II-73	89
II-74	90
II-75	91
II-76	92
II-77	93
II-78	94
II-79	95
II-80	96
II-81	97
II-82	98
II-84	99
II-87	100
II-88	101
II-92	102
II-93	103
II-94	104
II-96	105
III-01	106
III-02	107
III-03	108
III-06	109
III-07	110
III-08	111
III-09	112
III-11	113
III-12	114
III-13	115
III-18	116
III-20	401
III-21	117
III-22	118
III-23	119
III-24	120
III-25	121
III-26	122
III-27	123
III-28	124
III-29	125
III-31	126
III-32	127
III-33	128
III-34	129
III-35	130
III-39	131
III-40	132

Clean Copy

III-42	133
III-43	500
III-44	134
III-45	135
III-46	136
III-47	137
III-48	138
III-49	139
III-50	140
III-52	141
III-53	142
III-55	143
III-56	144
III-57	145
III-58	146
III-59	147
III-61	148
III-62	149
III-63	150
III-64	151
III-66	152
III-67	153
III-70	154
III-74	155
III-76	156
III-78	157
III-80	158
III-81	159
III-82	451
III-83	160
III-85	161
III-86	162
III-88	163 & 164
III-89	165
III-92	452
III-93	166
III-94	167
III-95	168
IV-04	273
IV-13	274
IV-14	275
IV-15	402
IV-17	276
IV-23	454
IV-26	403
IV-28	277
IV-31	278
IV-32	279
IV-35	455
IV-37	497
IV-38	280
IV-40	281
IV-42	282
IV-43	441
IV-44	283
IV-47	284
IV-53	498
IV-55	285
IV-61	286
IV-64	287
IV-65	288
IV-69	4
IV-72	289
IV-73	290
IV-80	291
IV-85	292
IV-93	293
TV-95	294
IV-96	295
IX-10	314
IX-13	315
IX-24	316
IX-38	317

IX-39	318
IX-48	319
IX-50	320
IX-56	321
IX-62	322
IX-65	323
IX-72	324
IX-77	325
IX-91	326
IX-96	327
V-01	458
V-03	296
V-04	297
V-07	298
V-08	299
V-09	300
V-11	404
V1-16	344
V1-19	345
V-12	301
V-17	459
V-20	302
V-24	303
V-25	460
V-28	405
V-35	461
V-38	406
V-39	389
V-40	304
V-41	305
V-47	463
V-48	306
V-49	464
V-55	499
V-57	307
V-58	465
V-61	308
V-64	309
V-68	484
V-71	496
V-74	310
V-75	467
V-80	311
V-81	312
V-87	313
V-90	468
VI-12	341
VI-13	342
VI-14	343
VI-16	344
VI-23	347
VI-24	348
VI-32	351
VI-39	352
VI-43	471
VI-44	409
VI-45	353
VI-49	501
VI-50	356
VI-53	357
VI-55	359
VI-58	361
VI-66	363
VI-67	364
VI-70	2
VI-71	472
VI-74	365
VI-75	366
VI-76	367
VI-77	3
VI-79	473
VI-80	368
VI-85	369

VI-87	370
VI-88	371
VI-90	474
VI-93	475
VI-95	374
VI-96	476
VII-17	169
VII-18	170
VII-19	171
VII-20	172
VII-21	173
VII-22	174
VII-23	175
VII-24	176
VII-25	480
VII-26	5
VII-27	177
VII-29	178
VII-32	179
VII-33	180
VII-35	181
VII-36	182
VII-39	183
VII-40	184
VII-41	185
VII-42	186
VII-43	187
VII-44	188
VII-45	189
VII-46	190
VII-47	415
VII-49	191
VII-50	192
VII-52	193
VII-53	194
VII-54	195
VII-55	196
VII-57	197
VII-58	198
VII-59	199
VII-62	200
VII-63	201
VII-64	202
VII-65	203
VII-66	204
VII-67	481
VII-71	205
VII-72	206
VII-73	207
VII-74	208
VII-76	209
VII-80	210
VII-81	211
VII-82	212
VII-84	213
VII-86	487
VII-87	214
VII-89	215
VII-90	216
VII-91	217
VII-92	218
VII-93	219
VII-96	220
VIII-09	221
VIII-10	222
VIII-12	223
VIII-13	224
VIII-16	225
VIII-17	226
VIII-18	227
VIII-19	228
VIII-20	229
VIII-21	230

Clean Copy

VIII-23	231
VIII-24	232
VIII-25	233
VIII-26	489
VIII-27	234
VIII-28	235
VIII-29	236
VIII-30	237
VIII-31	238
VIII-32	239
VIII-33	240
VIII-36	241
VIII-37	242
VIII-38	243
VIII-40	244
VIII-41	245
VIII-42	246
VIII-43	247
VIII-45	248
VIII-46	249
VIII-47	250
VIII-48	251
VIII-50	252
VIII-51	253
VIII-53	254
VIII-54	255
VIII-55	256
VIII-56	257
VIII-57	258
VIII-59	259
VIII-60	260
VIII-64	261
VIII-66	262
VIII-67	263
VIII-70	264
VIII-71	265
VIII-72	266
VIII-73	267
VIII-74	268
VIII-75	269
VIII-76	270
VIII-77	271
VIII-80	272
X-07	328
X-15	329
X-20	330
X-29	331
X-34	332
X-46	333
X-54	334
X-56	335
X-68	421
X-72	336
X-73	422
X-94	337
XI-13	423
XI-37	490
XI-43	424
XI-67	425
XI-81	426
XII-07	427
XII-35	428
XII-36	429
XII-59	430
XII-65	381
XII-92	431
XIII-03	375
XIII-04	432
XIII-19	433
XIII-24	376
XIII-51	377
XIII-52	378
XIII-67	379

Clean Copy

XIII-69	380
XIII-88	434
XIII-92	435
XV-22	388
XV-25	436
XV-62	437
XV-64	390
XV-84	391
XVI-19	438
XVI-36	382
XVI-53	439
XVI-60	383
XVI-66	384
XVI-74	385
XVI-76	386
XVI-77	387
XVII-31	392
XVII-40	440
XVII-48	393
XVII-76	394
XVII-87	395
XVII-95	396

Table 2a

List of informative probes for diagnosis of breast cancer

Clone ID	SEQ ID NO. in Sequence Listing
I-24	11
I-28	13
I-30	398
I-54	399
II-41	60
II-70	86
II-87	100
III-06	109
III-20	401
III-40	132
III-57	145
III-61	148
III-89	165
IV-14	275
IV-15	402
IV-26	403
IV-32	279
IV-53	498
IV-69	4
IV-80	291
IX-10	314
IX-38	317
IX-48	319
IX-77	325
V-11	404
V-55	499
V-80	311
VI-07	1
VI-48	355
VI-55	359
VI-70	2
VII-03	411
VII-15	414
VII-32	179
VII-39	183
VII-47	415
VII-48	416
VII-73	207
VII-77	418
VII-90	216
VIII-20	229
VIII-29	236
VIII-30	237
VIII-31	238
VIII-46	249
VIII-48	251
VIII-66	262
VIII-76	270
X-07	328
X-15	329
X-29	331
X-54	334
X-56	335
X-68	421
X-72	336
X-94	337
XI-13	423
XI-81	426
XII-07	427
XII-35	428

Clean Copy

Clone ID	Sequence ID
XII-59	430
XIII-19	433
XIII-52	378
XIII-92	435
XV-22	388
XV-25	436
XVI-36	382
XVI-53	439
XVI-66	384
XVI-76	386
XVI-77	387
XVII-31	392
XVII-40	440
XVII-48	393
XVII-76	394
XVII-87	395
XVII-95	396

Clean Copy

Table 2b

List of sequences of probes informative for breast cancer

Clone ID	SEQ ID NO. in Sequence Listing
I-13	444
I-14	397
I-24	11
I-25	12
I-28	13
I-30	398
I-37	482
I-42	445
I-48	19
I-54	399
I-60	25
I-72	446
I-81	31
I-82	32
I-86	447
I-88	400
I-95	448
II-02	33
II-03	34
II-06	36
II-07	37
II-10	39
II-21	45
II-23	46
II-24	47
II-25	48
II-27	50
II-33	55
II-34	56
II-41	60
II-42	61
II-46	64
II-47	449
II-48	66
II-52	68

Clean Copy

II-57	73
II-58	74
II-59	75
II-60	76
II-61	77
II-62	78
II-64	80
II-67	83
II-69	85
II-70	86
II-74	90
II-80	96
II-82	98
II-84	99
II-87	100
II-88	101
II-96	105
III-01	106
III-02	107
III-06	109
III-08	111
III-12	114
III-13	115
III-17	450
III-18	116
III-20	401
III-21	117
III-23	119
III-24	120
III-25	121
III-26	122
III-27	123
III-28	124
III-29	125
III-32	127
III-33	128
III-35	130
III-39	131
III-40	132
III-42	133
III-45	135
III-46	136
III-47	137
III-48	138
III-56	144
III-57	145
III-58	146

Clean Copy

III-59	147
III-61	148
III-62	149
III-63	150
III-64	151
III-66	152
III-67	153
III-70	154
III-74	155
III-75	156
III-78	157
III-80	158
III-81	159
III-82	451
III-85	161
III-86	162
III-88	163 + 164
III-89	165
III-92	452
III-93	166
III-95	168
III-96	452
IV-04	273
IV-13	274
IV-14	275
IV-15	402
IV-17	276
IV-23	454
IV-26	403
IV-31	278
IV-32	279
IV-35	455
IV-37	497
IV-38	280
IV-42	282
IV-43	441
IV-47	284
IV-53	498
IV-61	286
IV-64	287
IV-69	4
IV-72	289
IV-80	291
IV-85	292
IV-93	457
IV-96	295
IX-10	314
IX-13	315

Clean Copy

IX-24	316
IX-38	317
IX-39	318
IX-48	319
IX-50	320
IX-56	321
IX-62	322
IX-65	323
IX-72	324
IX-77	325
IX-91	326
IX-96	327
V-01	458
V-03	296
V-04	297
V-07	298
V-08	299
V-11	404
V-12	301
V-17	459
V-24	303
V-25	460
V-28	405
V-38	461
V-38	406
V-39	389
V-41	305
V-47	463
V-49	464
V-55	499
V-57	307
V-58	465
V-61	308
V-64	309
V-65	466
V-68	484
V-71	496
V-74	310
V-75	467
V-80	311
V-90	468
VI-03	338
VI-04	339
VI-07	1
VI-08	340
VI-09	469
VI-12	341
VI-13	342

Clean Copy

VI-14	343
VI-16	344
VI-19	345
VI-20	346
VI-21	470
VI-23	347
VI-24	348
VI-25	408
VI-26	349
VI-32	351
VI-39	352
VI-43	471
VI-44	409
VI-45	353
VI-48	355
VI-49	501
VI-50	356
VI-53	357
VI-55	359
VI-58	361
VI-66	363
VI-67	364
VI-70	2
VI-71	472
VI-74	365
VI-75	366
VI-76	367
VI-77	3
VI-79	473
VI-80	368
VI-85	369
VI-87	370
VI-88	371
VI-90	474
VI-93	475
VI-95	374
VI-96	476
VII-02	410
VII-03	411
VII-06	477
VII-08	412
VII-09	413
VII-10	478
VII-11	479
VII-15	414
VII-17	169
VII-19	171
VII-21	173

Clean Copy

VII-22	174
VII-23	175
VII-24	176
VII-25	480
VII-26	5
VII-27	177
VII-29	178
VII-32	179
VII-33	180
VII-36	182
VII-39	183
VII-41	185
VII-42	186
VII-43	187
VII-46	190
VII-47	415
VII-48	416
VII-49	191
VII-54	195
VII-57	197
VII-58	198
VII-59	199
VII-62	200
VII-63	417
VII-64	202
VII-66	204
VII-67	481
VII-72	206
VII-73	207
VII-77	418
VII-80	210
VII-82	212
VII-86	487
VII-87	214
VII-90	216
VII-91	217
VII-92	218
VII-93	219
VII-96	220
VIII-09	221
VIII-10	222
VIII-13	224
VIII-16	225
VIII-20	229
VIII-21	230
VIII-23	231
VIII-24	232
VIII-25	233

Clean Copy

VIII-26	489
VIII-27	234
VIII-28	235
VIII-29	236
VIII-30	237
VIII-31	238
VIII-32	239
VIII-33	240
VIII-34	419
VIII-38	243
VIII-40	244
VIII-41	245
VIII-46	249
VIII-48	251
VIII-55	256
VIII-57	258
VIII-59	259
VIII-60	260
VIII-61	420
VIII-64	261
VIII-66	262
VIII-73	267
VIII-74	268
VIII-76	270
VIII-80	272
X-07	328
X-15	329
X-20	330
X-29	331
X-34	332
X-46	333
X-54	334
X-56	335
X-68	421
X-72	336
X-73	422
X-94	337
XI-13	423
XI-37	490
XI-43	424
XI-67	425
XI-81	426
XII-07	427
XII-35	428
XII-36	429
XII-59	430
XII-65	381
XII-92	431

Clean Copy

XIII-03	375
XIII-04	432
XIII-19	433
XIII-24	376
XIII-51	377
XIII-52	378
XIII-67	379
XIII-69	380
XIII-88	434
XIII-92	435
XV-22	388
XV-25	436
XV-62	437
XV-64	390
XV-84	391
XVI-19	438
XVI-36	382
XVI-53	439
XVI-60	383
XVI-66	384
XVI-74	385
XVI-76	386
XVI-77	387
XVII-31	392
XVII-40	440
XVII-48	393
XVII-76	394
XVII-87	395
XVII-95	396

Clean Copy

Table 3

List of informative probes (Clone ID) selected for breast cancer diagnosis based on their occurrence criterion during variable selection

Occurrence*	Clone ID
100%	XVI-66,VIII-66,VII-03,XIII-19,XII-35,,IV-53,I-30,III-06,XV-22,VII-15,VII-39,IX-39,III-40,VII-32
90%	V-11,XIII-92,VIII-29,XVI-53,XVI-77,XI-13,IV-14,V-80,VII-48,
80%	XIII-52,VIII-30,IX-38
70%	X-29,VIII-48
60%	IX-10,X-68,VII-77
50%	IV-15
40%	II-70,V-55
20%	XVI-36,III-61,IV-69,X-72
10%	IX-77,X-94
5%	XII-59,XVI-76,I-54,X-54,VI-07,VII-47,XVII-31,XVII-87,XVII-48
In at least one model	II-41,III-57,III-89,VII-73,XV-25,IV-26,VII-90,VIII-20,I-28,VIII-76,III-20,XVII-76,VIII-46,VI-70,VIII-31,II-87,VI-55,X-07,X-15,XII-07,XVII-95,I-24,IV-32,VI-48,IV-80,IX-48,X-56,XVII-40

*100% = Genes appearing in all the 75 cross validated models; 90% = Additional genes appearing in at least 68 out of 75 cross validated models; 5% = Additional genes appearing in at least 4 out of 75 cross validated models and so on.

Clean Copy

Table 4a

List of informative probes for diagnosis of Alzheimer disease

Clone ID	SEQ ID NO. in Sequence Listing
I-34	15
I-58	24
II-03	34
II-05	35
II-06	36
II-10	39
II-24	47
II-25	48
II-26	49
II-33	55
II-34	56
II-42	61
II-57	73
II-61	77
II-69	85
II-75	91
II-84	99
II-88	101
II-94	104
III-02	107
III-06	109
III-08	111
III-13	115
III-23	119
III-26	122
III-35	130
III-39	131
III-43	500
III-44	134
III-53	142
III-56	144
III-63	150
III-74	155
III-80	158
III-85	161
IV-31	278
IV-80	291
V-03	296
V-04	297
V-07	298
V-12	301
V-80	311
VI-04	339
VI-12	341
VI-14	343
VI-20	346
VI-23	347
VI-48	355

Clone ID	SEQ ID NO. in Sequence Listing
VI-50	356
VI-53	357
VI-74	365
VI-76	367
VI-87	370
VI-88	371
VI-95	374
VII-19	171
VII-21	173
VII-36	182
VII-42	186
VII-43	187
VII-46	190
VII-59	199
VII-63	201
VII-66	204
VII-72	206
VII-73	207
VI-12	344
VI-14	345
VII-91	217
VII-93	219
VIII-09	221
VIII-28	235
VIII-30	237
VIII-32	239
VIII-33	240
VIII-41	245
VIII-42	246
VIII-48	251
VIII-64	261
VIII-67	263

Clean Copy

Table 4b

List of sequences of probes informative for Alzheimer disease

Clone ID	SEQ ID NO. in Sequence Listing
I-10	6
I-15	7
I-17	8
I-19	9
I-22	10
I-24	11
I-25	12
I-28	13
I-31	14
I-34	15
I-38	16
I-39	17
I-40	18
I-48	19
I-49	20
I-53	21
I-56	22
I-57	23
I-58	24
I-60	25
I-64	26
I-67	27
I-69	28
I-77	29
I-80	30
I-81	31
I-82	32
II-02	33
II-03	34
II-05	35
II-06	36
II-07	37
II-08	38
II-10	39
II-11	40
II-12	41
II-13	42
II-15	43
II-16	44
II-21	45
II-23	46
II-24	47
II-25	48
II-26	49
II-27	50

Clean Copy

II-29	51
II-30	52
II-31	53
II-32	54
II-33	55
II-34	56
II-38	57
II-39	58
II-40	59
II-41	60
II-42	61
II-43	62
II-44	63
II-46	64
II-47	65
II-48	66
II-50	67
II-52	68
II-53	69
II-54	70
II-55	71
II-56	72
II-57	73
II-58	74
II-59	75
II-60	76
II-61	77
II-62	78
II-63	79
II-64	80
II-65	81
II-66	82
II-67	83
II-68	84
II-69	85
II-70	86
II-71	87
II-72	88
II-73	89
II-74	90
II-75	91
II-76	92
II-77	93
II-78	94
II-79	95
II-80	96
II-81	97
II-82	98
II-84	99
II-87	100
II-88	101

Clean Copy

II-92	102
II-93	103
II-94	104
II-96	105
III-01	106
III-02	107
III-03	108
III-06	109
III-07	110
III-08	111
III-09	112
III-11	113
III-12	114
III-13	115
III-21	117
III-22	118
III-23	119
III-24	120
III-25	121
III-26	122
III-27	123
III-28	124
III-29	125
III-31	126
III-32	127
III-33	128
III-34	129
III-35	130
III-39	131
III-40	132
III-42	133
III-43	500
III-44	134
III-45	135
III-46	136
III-47	137
III-48	138
III-49	139
III-50	140
III-52	141
III-53	142
III-55	143
III-56	144
III-57	145
III-58	146
III-59	147
III-61	148
III-62	149
III-63	150
III-64	151
III-66	152

Clean Copy

III-67	153
III-70	154
III-74	155
III-75	156
III-78	157
III-80	158
III-81	159
III-83	160
III-85	161
III-86	152
III-88	163/164
III-89	165
III-93	166
III-94	167
III-95	168
VII-17	169
VII-18	170
VII-19	171
VII-20	172
VII-21	173
VII-22	174
VII-23	175
VII-24	176
VII-27	177
VII-29	178
VII-32	179
VII-33	180
VII-35	181
VII-36	182
VII-39	183
VII-40	184
VII-41	185
VII-42	186
VII-43	187
VII-44	188
VII-45	189
VII-46	190
VII-49	191
VII-50	192
VII-52	193
VII-53	194
VII-54	195
VII-55	196
VII-57	197
VII-58	198
VII-59	199
VII-62	200
VII-63	201
VII-64	202
VII-65	203
VII-66	204

Clean Copy

VII-71	205
VII-72	206
VII-73	207
VII-74	208
VII-76	209
VII-80	210
VII-81	211
VII-82	212
VII-84	213
VII-87	214
VII-89	215
VII-90	216
VII-91	217
VII-92	218
VII-93	219
VII-96	220
VIII-09	221
VII-10	222
VII-12	223
VII-13	224
VII-16	225
VII-17	226
VII-18	227
VII-19	228
VII-20	229
VII-21	230
VII-23	231
VII-24	232
VII-25	233
VII-28	235
VII-29	236
VII-30	237
VII-31	238
VII-32	239
VII-33	240
VII-36	241
VII-37	242
VII-38	243
VII-40	244
VII-41	245
VII-42	246
VII-43	247
VII-45	248
VII-46	249
VII-47	250
VII-48	251
VII-50	252
VII-51	253
VII-53	254
VII-54	255
VII-55	256

Clean Copy

VII-56	257
VII-57	258
VII-59	259
VII-60	260
VII-64	261
VII-66	262
VII-67	263
VII-70	264
VII-71	265
VII-72	266
VII-73	267
VII-74	268
VII-75	269
VII-76	270
VII-77	271
VII-80	272
IV-04	273
IV-13	274
IV-14	275
IV-17	276
IV-28	277
IV-31	278
IV-32	279
IV-38	280
IV-40	281
IV-42	282
IV-44	283
IV-47	284
IV-55	285
IV-61	286
IV-64	287
IV-65	288
IV-72	289
IV-73	290
IV-80	291
IV-85	292
IV-93	293
IV-95	294
IV-96	295
V-03	296
V-04	297
V-07	298
V-08	299
V-09	300
V-12	301
V-20	302
V-24	303
V-40	304
V-41	305
V-48	306
V-57	307

Clean Copy

V-61	308
V-64	309
V-74	310
V-80	311
V-81	312
V-87	313
VI-13	342
VI-14	343
VI-16	344
VI-23	347
VI-24	348
VI-28	350
VI-32	351
VI-39	352
VI-45	353
VI-46	354
VI-49	501
VI-50	356
VI-53	357
VI-54	358
VI-55	359
VI-57	360
VI-58	361
VI-63	362
VI-66	363
VI-67	364
VI-74	365
VI-75	366
VI-76	367
VI-80	368
VI-85	369
VI-87	370
VI-88	371
VI-91	372
VI-94	373
VI-95	374
I-14	397
I-30	398
I-54	399
I-88	400
III-20	401
IV-15	402
IV-26	403
V-11	404
IV-28	405
IV-38	406
IV-45	407
VI-44	409
VII-47	415
I-42	445
I-86	447

Clean Copy

I-95	448
III-82	451
III-92	452
IV-23	454
IV-35	455
IV-82	456
V-01	458
V-17	459
V-25	460
V-35	461
V-42	462
V-47	463
V-49	464
V-58	465
V-75	467
V-90	468
VI-43	471
VI-71	472
VI-79	473
VI-90	474
VI-93	475
VII-25	480
VII-67	481
I-37	482
V-52	483
V-68	484
V-92	485
VI-42	486
VII-86	487
VII-88	488
IV-29	491
V-15	491
V-39	493
V-54	494
V-59	495
V-71	496

Clean Copy

Table 5

Samples

Diagnosis	No. of women
Normal/Benign	42*
DCIS	3
Invasive cancer	26

*From one woman, whole blood was collected at weeks 1,2,3,4,5 following menstruation. Hence, the number of unique normal/benign samples tested in the experiment is 75.

Information about women with breast cancer

Sample	*AGE*	Stage	*Cancer type*	*Size hist.* (mm)	*Nodes*
1	51	II	IDC	20	1/7
2	84	II	IDC	22	2/2
3	50	I	DCIS+ 1 IDC	>50 DCIS; 5 x 14	0/7
4	47	I	IDC	15	0
5	69	III	ILC g.2 + tubular adenocarcinoma	50 + 3	1 av 12 + 1 av 7
6	50	II	IDC	24	0
7	65	I	IDC	15	0
8	63	II	IDC	23	0
9	55	I	IDC + DCIS	4	0 av 1
10	52	0	DCIS + small colloid carcinoma foci	50 + 3	0
11	60	II	IDC	24	0
12	54	I	IDC	11	0
13		0	DCIS	20	0
14	49	0	DCIS	9	0
15	48	I	IDC	4	0
16	56	I	IDC	4	0
17	68	I	IDC	14	0
18	68	I	IDC	7	0
19	63	I	IDC	10	0
20	45	I	IDC	19	1
21	57	III	IDC	60	8/20

22	55	II	IDC/DCIS	35 + 55	0
23	71	I	IDC/extensive DCIS	8	0
24	56	I	IDC	9	?
25	66	II	IDC	26	0
26	66	I	IDC	15	?
27	61	I	IDC	9	?
28	?	?	?	?	?
29	65	I	IDC	11	0

Other diseases/conditions present in the women tested

Other diseases/conditions present in the women tested

Disease/condition
Diabetes
Asthma
Ulcerous colitis
Hemochromatose
Crohn's disease
Fibromyalgia
Psoraiasis
Atopic eczema
Rheumatism
Allergies

Prior history of cancer in the women tested

Cancer type	No. of women
Breast	3
Colon	2
Stomach	1
Skin	1

Clean Copy

Table 6

Number of samples tested by double cross validation and success of the diagnostic test for breast cancer based on selected ionformative genes

Number of samples tested by double cross validation

Number of unique samples tested	75
Number of unique non cancer samples tested	46
Number of cancer samples tested	29

Success of the diagnostic test for breast cancer based on selected informative genes

Occurrence in percentage*	Number of informative probes	Specificity	Sensitivity	Accuracy	False Positive rate	False negative rate	Total error rate
100.00	23	84.78	75.86	81.33	15.22	24.14	18.67
90.00	44	91.30	79.31	86.67	8.70	20.69	13.33
80.00	51	86.96	79.31	84.00	13.04	20.69	16.00
70.00	54	89.13	75.86	84.00	10.87	24.14	16.00
60.00	58	89.13	75.86	84.00	10.87	24.14	16.00
50.00	59	89.13	75.86	84.00	10.87	24.14	16.00
40.00	63	89.13	75.86	84.00	10.87	24.14	16.00
30.00	66	86.96	75.86	82.67	13.04	24.14	17.33
20.00	74	89.13	75.86	84.00	10.87	24.14	16.00
10.00	79	89.13	75.86	84.00	10.87	24.14	16.00
5.00	90	86.96	79.31	84.00	13.04	20.69	16.00
1.33	139	84.78	72.41	80.00	15.22	27.59	20.00

*100% = Genes appearing in all the 75 cross validated models; 90% = Genes appearing in at least 68 out of 75 cross validated models; 5% = Genes appearing in at least 4 out of 75 cross validated models; and so on.

Clean Copy

Table 7

Double cross-validation and details of the success of the diagnostic test for Alzheimer disease based on the expression 182 informative genes

Validation Result

Total number of samples tested	14
Number of Alzhelmer's disease samples tested	7
Number of Alzhelmer's disease samples incorrectly predicted	1
Number of non-Alzhelmer's disease samples tested	7
Number of non-Alzhelmer's disease samples incorrectly predicted	0

Success of diagnostic test

Performance	**Description**	**%**
Accuracy	Percentage of the total number of predictions that were correct	92.9
Sensitivity	Percentage of positive cases that were correctly identified	85.7
Specificity	Percentage of negatives cases that were correctly predicted	100
False positive rate	Percentage of negatives cases that were incorrectly classified as positive	0.0
False negative rate	Percentage of positive cases that were incorrectly classified as negative	14.3
Total error rate	Percentage of the total cases incorrectly predicted	7.1

Clean Copy

Table 8

Some relevant features of the blood donors. **B**, Female donors with breast cancer; **N**, Female donors with suspected mammogram but no breast cancer; **IDC**, invasive ductal carcinoma; **DCIS**, ductal carcinoma in situ; **na**, not available **nd**, not determined; ++, no degradation of mRNA and no ribosomal contamination in the sample, **+**, no degradation of mRNA but ribosomal contamination in the sample.

		AGE	Cancer type/ breast abnormality	Size Hist. (mm)	mRNA Quality
1	B1	na	IDC	5	++
2	B2	49	DCIS	8	nd
3	B3	54	IDC	18	++
4	B4	59	IDC	12	+
5	B5	61	DCIS + micro invasive cancer	15+1.5	++
6	B6	55	IDC	12+17	nd
7	B6		IDC	12+17	nd
8	N1	45	Fibroadenoma	-	nd
9	N2	52	na	-	+
10	N3	55	Cyst	-	++
11	N4	54	na	-	++
12	N5	51	Benign ductal epithelium	-	nd
13	N6	57	Benign	-	nd

14	N7	50	na	-	++
15	N8	52	na	-	+

Clean Copy

Table 9

List of sequences of probes informative for both alzheimer and breast cancer disease

Clone ID	SEQ ID NO. in Sequence Listing
I-24	11
I-25	12
I-28	13
I-48	19
I-60	25
I-81	31
I-82	32
II-02	33
II-03	34
II-06	36
II-07	37
II-10	39
II-21	45
II-23	46
II-24	47
II-25	48
II-27	50
II-33	55
II-34	56
II-41	60
II-42	61
II-46	64
II-47	65
II-48	66
II-52	68
II-57	73
II-58	74
II-59	75
II-60	76
II-61	77
II-62	78
II-64	80
II-67	83
II-69	85
II-70	86
II-74	90
II-80	96
II-82	98
II-84	99
II-87	100
II-88	101
II-96	105
III-01	106

Clean Copy

III-02	107
III-06	109
III-08	111
III-12	114
III-13	115
III-18	116
III-21	117
III-23	119
III-24	120
III-25	121
III-26	122
III-27	123
III-28	124
III-29	125
III-32	127
III-33	128
III-35	130
III-39	131
III-40	132
III-42	133
III-45	135
III-46	136
III-47	137
III-48	138
III-56	144
III-57	145
III-58	146
III-59	147
III-61	148
III-62	149
III-63	150
III-64	151
III-66	152
III-67	153
III-70	154
III-74	155
III-5	156
III-78	157
III-80	158
III-81	159
III-85	161
III-86	162
III-88	163/164
III-89	165
III-93	166
III-95	168
IV-04	273

Clean Copy

IV-13	274
IV-14	275
IV-17	276
IV-31	278
IV-32	279
IV-38	280
IV-42	282
IV-47	284
IV-61	286
IV-64	287
IV-72	289
IV-80	291
IV-85	292
IV-93	293
IV-96	295
V-03	296
V-04	297
V-07	298
V-08	299
V-12	301
V-24	303
V-41	305
V-57	307
V-61	308
V-64	309
V-74	310
V-80	311
VI-12	341
VI-14	343
VI-23	347
VI-50	356
VI-53	357
VI-74	365
VI-76	367
VI-87	370
VI-88	371
VI-95	374
VII-19	171
VII-21	173
VII-22	174
VII-23	175
VII-24	176
VII-27	177
VII-29	178
VII-32	179
VII-33	180
VII-36	182

Clean Copy

Clean Copy

VIII-60	260
VIII-64	261
VIII-66	262
VIII-73	267
VIII-74	268
VIII-76	270
VIII-80	272